Journal of Statistical Planning and Inference 155 (2014) 117–126

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Design issues related to allocation of experimental units with known covariates into two treatment groups

Samrat Hore [a], Anup Dewanji [b], Aditya Chatterjee [c,*]

[a] Department of Statistics, Tripura University, India
[b] Applied Statistics Unit, Indian Statistical Institute, Kolkata, India
[c] Department of Statistics, University of Calcutta, India

ARTICLE INFO

Article history:
Received 14 May 2013
Received in revised form 28 February 2014
Accepted 19 June 2014
Available online 9 July 2014

Keywords:
D-optimality
A-optimality
D_s-optimality
A_s-optimality
Allocation design
Near-optimal design
Variable neighborhood search
Robust allocation design

ABSTRACT

An experimental situation corresponding to an ANCOVA model involving several known covariates is of interest in many practical situations related to various disciplines. The design problem here is to determine an optimal allocation rule for the experimental units with known covariate(s) into two treatments such that both treatment and covariate effects, or only treatment effect, are efficiently estimated in terms of the commonly used *D*- or *A*-optimalities. The present work deals with development of a new allocation algorithm for this purpose. The algorithm generates efficient allocation in comparison to random allocation and allocation designs already reported in the literature. At the same time the algorithm requires less computational work. With large number of experimental units and several covariates, the exact optimal allocation is computationally intractable. In such situation, we have proposed an algorithm for finding a near-optimal allocation design that is motivated by computationally tractable optimization technique. Through simulation experiments and real life data analysis, we have demonstrated the efficiency of the proposed algorithm.

1. Introduction

The problem of comparing several treatments in presence of a single or multiple covariates has been of interest in many application areas including chemical industry, agriculture, biology and medicine (Harville, 1974; Rao, 1973; Hinkelmann and Kempthorne, 2005, 2007; Saville and Wood, 1991; Cartwright et al., 1968; Kalbfleisch and Prentice, 2002; Hu and Hu, 2012). Generally, in such problems, we consider a linear model such as in *ANCOVA*. The usual design issue is to find the optimum values of the covariates with respect to the well-known *D*- or *A*-optimalities (Kiefer, 1959; Das et al., 2003; Dey and Mukerjee, 2006). Another type of design issue is concerned with the allocation of several treatments to a fixed number experimental units with known covariates, so that the treatment effect or the covariate effect or both can be efficiently estimated with respect to some desired optimality criteria, like *D*- or *A*-optimalities. Such design problem may be relevant in different areas like veterinary science (Saville and Wood, 1991), engineering science (Harville, 1974) and medical science (Kalbfleisch and Prentice, 2002) as well.

* Corresponding author. Tel.: +91 9433015528; fax: +91 3324614849.
E-mail addresses: acbustat1@gmail.com, ac_bustat1@yahoo.co.in (A. Chatterjee).

http://dx.doi.org/10.1016/j.jspi.2014.06.002
0378-3758/

In the present paper we address this allocation design problem in which, as also mentioned in Harville (1974), owing to the large number of possible allocations, finding the exact *D*- or *A*-optimal allocation design is generally not practical. For example, with n experimental units to be allocated into two treatments, one needs to examine $(2^{n-1}-1)$ such non-trivial allocation schemes for the optimal solution making the procedure to be exponentially complex. However, as we shall explain in later sections, the actual search may be judiciously confined to a smaller number of schemes leading to an optimal or near-optimal solution.

In this regard, an efficient allocation algorithm through a multistage procedure has been proposed by Harville (1974) that allocates one or more units at each stage and improvement in the allocation scheme is achieved through an iterative algorithm that induces small changes in the design at each iteration. An improvement on his initial allocation is obtained by switching the allocation of one experimental unit from one treatment to another treatment, or by exchanging the allocation of experimental units from one treatment to the other. Cook and Nachstsheim (1989) considered an extension of Harville's problem by incorporating a block structure (with pre-specified block sizes) in the design and obtained nearly *D*-optimal design by carrying out essentially the same method as advocated by Harville (1974). In the present paper, the two allocation schemes as mentioned in Harville (1974) have been critically examined with reference to several allocation designs *viz.* the randomized allocation or some optimal or near-optimal allocation schemes introduced here.

In Section 2, the problem has been formulated in the *ANCOVA* framework. A quick method has been proposed in Section 3 to obtain an initial design. This has been found to be nearly fully efficient for single covariate when the number of experimental units is small. The proposed algorithm has been described in Section 4. A compromise between *D*- and *A*-optimalities has been proposed in Section 5 which has been termed as *robust* optimality and the proposed algorithm has been extended in this context as well. A detailed simulation study under various univariate and multivariate probability models for the covariates are considered in Section 6. The efficacy of the proposed algorithm in comparison to completely randomized allocation and other available schemes, under various setups, has been established through this simulation study. Two real life examples have also been considered with data from animal health laboratory study (Saville and Wood, 1991) and clinical trial on patients with severe aplastic anemia (Kalbfleisch and Prentice, 2002). Section 7 ends with some concluding remarks and scope of further research.

2. Problem formulation

Consider the one-way ANCOVA model with two treatments and p covariates whose values are assumed to be known for the n experimental units. Specifically, the linear model used for the analysis is given by

$$y_{li} = \mu_l + \sum_{j=1}^{p} \beta_j x_{lij} + \epsilon_{li}, \quad \text{for } i = 1(1)n_l,\ l = 1, 2, \tag{2.1}$$

where n_l (≥ 1) denotes the number of times the lth treatment is replicated and $\sum_{l=1}^{2} n_l = n$. Here, y_{li}, μ_l, β_j, x_{lij} and ϵ_{li} respectively denote the response from ith experimental unit receiving the lth treatment, the lth treatment effect, the regression coefficient for the jth covariate, the value of the jth covariate corresponding to ith experimental unit receiving the lth treatment, and the observational error assumed to be *iid* following $N(0, \sigma^2)$ distribution.

With reference to the model (2.1), the parameters of interest are (μ_1, μ_2) and $(\beta_1, \beta_2, \ldots, \beta_p)$. Write $\theta = (\mu_1, \mu_2, \beta_1, \beta_2, \ldots, \beta_p)$. For estimability of θ, it is necessary that $n \geq (p+2)$. The optimality criteria depends on the information matrix I given by

$$I = \begin{pmatrix} n_1 & 0 & \mathbf{x}_{\bullet 1}^T \\ 0 & n_2 & \mathbf{x}_{\bullet 2}^T \\ \mathbf{x}_{\bullet 1} & \mathbf{x}_{\bullet 2} & \mathbf{x}^T\mathbf{x} \end{pmatrix},$$

where $\mathbf{x}_{\bullet l}^T = 1_{n_l}^T \mathbf{x}_l$ with 1_{n_l} being the $(n_l \times 1)$ vector of all ones, $\mathbf{x}_l$ the $(n_l \times p)$ matrix given by $\mathbf{x}_l = ((x_{lij}))$, for $l = 1, 2$, and $\mathbf{x}$ is the $(n \times p)$ matrix given by $\mathbf{x}^T = (\mathbf{x}_1^T, \mathbf{x}_2^T)$. For *D*- and *A*-optimalities (for efficient estimation of both treatment and covariate effects), we need to minimize $\det(I^{-1})$ and $\text{trace}(I^{-1})$, respectively. Consider the partition of I as

$$I = \begin{pmatrix} A & B \\ B^T & D \end{pmatrix},$$

where $A = \text{Diag}(n_1, n_2)$; $B^T = (\mathbf{x}_{\bullet 1}, \mathbf{x}_{\bullet 2})$ and $D = \mathbf{x}^T\mathbf{x}$. Following Cook and Nachstsheim (1989), D_s- and A_s-optimal designs correspond to minimization of determinant or trace, respectively, of the covariance matrix of the least squares estimates of a selected subset of parameters under investigation. We investigate D_s- (A_s-)optimal designs that minimize the determinant (trace) of the covariance matrix of the least square estimates of only the treatment effect. Here, we need to consider the (1, 1) block entry of the inverse of the partitioned matrix (I). By following Rao and Bhimasankaran (2000, p. 138), this entry is given by

$$I^{*-1} = A^{-1} + A^{-1}B(D - B^TA^{-1}B)^{-1}B^TA^{-1}. \tag{2.2}$$